

A GENERAL CLASS OF MULTINOMIAL MIXTURE MODELS FOR ANURAN CALLING SURVEY DATA

J. ANDREW ROYLE¹ AND WILLIAM A. LINK

USGS Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, Maryland 20708 USA

Abstract. We propose a general framework for modeling anuran abundance using data collected from commonly used calling surveys. The data generated from calling surveys are indices of calling intensity (vocalization of males) that do not have a precise link to actual population size and are sensitive to factors that influence anuran behavior. We formulate a model for calling-index data in terms of the maximum potential calling index that could be observed at a site (the “latent abundance class”), given its underlying breeding population, and we focus attention on estimating the distribution of this latent abundance class. A critical consideration in estimating the latent structure is imperfect detection, which causes the observed abundance index to be less than or equal to the latent abundance class. We specify a multinomial sampling model for the observed abundance index that is conditional on the latent abundance class. Estimation of the latent abundance class distribution is based on the marginal likelihood of the index data, having integrated over the latent class distribution. We apply the proposed modeling framework to data collected as part of the North American Amphibian Monitoring Program (NAAMP).

Key words: *abundance index; amphibian monitoring; anurans; detection probability; frogs; latent class model; NAAMP; population index; site occupancy.*

INTRODUCTION

In recent years, there has been a dramatic increase of interest in the status of amphibian populations, owing to widespread declines in many species, across taxa (Alford and Richards 1999, Houlahan et al. 2000, Reaser 2000). As a result, state and federal agencies and other organizations have established amphibian monitoring initiatives. These include the Amphibian Research and Monitoring Initiative (ARMI; Hall and Langtimm 2001) initiated by the U.S. Department of Interior, and the North American Amphibian Monitoring Program (NAAMP; Weir and Mossman 2005). NAAMP is perhaps the most geographically extensive program, focusing primarily on monitoring anuran (i.e., frog and toad) populations throughout North America.

Because it is difficult or impossible to count individuals, most anuran monitoring efforts collect ordinal data, with levels representing the intensity of vocalization, which are used as an index to population size. We focus our discussion and model development on data collected as part of NAAMP, noting that the methods are easily adapted to other surveys producing ordered categorical data. The NAAMP index takes values $y = 0, 1, 2, 3$; indicating no frogs heard ($y = 0$); discrete, nonoverlapping calls ($y = 1$); discrete, overlapping calls ($y = 2$); and a full chorus of continuous, overlapping calls ($y = 3$) (Weir and Mossman 2005). Many other state and provincial monitoring programs

use this or similar indices. NAAMP surveys consist of three visits (temporal replicates) to each of a large number of sites made over the breeding period of all (vocal) anurans in the geographic stratum where samples are located. The data generated from these calling surveys consist of site-specific index histories, say y_i . Typical data for a single species might therefore be $y = (0, 2, 0)$ for three replicate samples at a single site. In this paper, we develop models for spatially and temporally replicated calling survey data of this sort.

Ideally, anuran monitoring programs would give exact population sizes; however, this is nearly impossible. A more reasonable goal, although still fraught with difficulties, is to produce indices that could be used to monitor and evaluate temporal patterns of change in population size. In this regard, there are two important problems with calling-index data that must be considered. First, there is not a precise link between calling-index values and population size. That is, the index values do not correspond to a well-defined number or even range of numbers of individuals. This is in contrast to simple counts of organisms that can be interpreted as realizations of a binomial random variable and for which various estimators of abundance can be calculated (e.g., Carroll and Lombard 1985, Royle 2004a). We suggest, however, that the precise relation between calling-index values and population size need not be specified for calling-index values to be useful. Instead, we propose a model for the calling-index data at a site that is parameterized in terms of the site's maximum supportable calling index, a latent feature clearly associated with population size, and a feature

Manuscript received 29 November 2004; accepted 31 January 2004. Corresponding Editor: S. R. Lele.

¹ E-mail: andy_royle@usgs.gov

that is estimable from calling-index data under replicated sampling. This defines a latent class model for the index data wherein estimation of the latent class structure becomes the focus of the analysis. The second problem with calling-index data is that detectability of anurans varies in response to many factors. Because the observed index value is based on calling activity (and not an actual count of frogs capable of calling), factors that influence anuran calling contribute to variation in detection probability across samples. Such factors include temperature, time of day, cloud cover, precipitation, wind conditions, and variation over time within a season (Bishop et al. 1997, Bridges and Dorcas 2000, Johnson and Batie 2001, Oseen and Wassersug 2002; L. A. Weir, J. A. Royle, P. Nanjappa, and R. E. Jung, *unpublished manuscript*). Variation in these factors across sample sites and visits can induce considerable variation in observed index histories. It is necessary to either control for these factors, or to model their effects so that objective and meaningful inferences about population status can be made. Our formulation of the model for the observed index data conditioned on the latent class membership of each site allows for modeling the effect of covariates that influence detection.

One approach for addressing these two issues more formally (MacKenzie et al. 2002) is to reduce the data to “observed presence/absence” (i.e., detection/non-detection), from which summaries of occurrence that account for variation in detectability can be estimated. While appealing for its simplicity, this approach represents only a crude characterization of the distribution of abundance across sites and will be insensitive to changes in abundance, especially for low or high abundance species. Royle (2004b) extended this notion to a model for the index data by viewing the observed index values as binomial counts. Although convenient, the binomial model is an obvious misspecification and overly restrictive because it is parameterized by a single detection probability parameter.

In this paper, we propose a general multinomial sampling model for ordinal categorical calling index data that are replicated in space and time. In *The multinomial mixture model*, we describe this model that is based on the view that each sample site possesses a “latent abundance class,” defined as the maximum index value that the population at that site could generate. The model consists of a component describing the latent class distribution and a state-dependent multinomial sampling distribution for the observed abundance index data at a site, the parameters of which govern the conditional distribution of the data given the latent abundance class. Construction of a large number of reduced models is possible under an alternative parameterization of the model given in *Alternative parameterization of the conditional sampling distributions*. In *Application to anuran calling survey data*, we provide an application of the proposed modeling framework to

data generated from NAAMP. We discuss the results and some topics of future research in *Discussion*.

THE MULTINOMIAL MIXTURE MODEL

The sampling design considered here specifies that R sites are sampled on T occasions during the breeding season (T may be allowed to vary across sites, and usually will, but for simplicity, we assume that T is constant for all sites). Let y_{it} be the observed index value at site i on sampling occasion t . We assume that site-specific populations are closed in the sense that the underlying breeding population is constant for all sampling occasions in the interval $t = 1$ to $t = T$.

Possible abundance classes (N) are determined by the calling-index definition. For the NAAMP index, N takes on four possible values, $N = 0, 1, 2, 3$, defining possible abundance classes that each site may belong to but that may not be observed during the sampling activity at that site. For example, a site with latent abundance class $N = 3$ could yield observed index values of $y = 0, 1, 2$, or 3 . A site with $N = 2$ could yield $y = 0, 1$, or 2 . Let N_i be the true abundance class at site i . N_i is the maximum index value that could be observed at site i , given its breeding population. We refer to N_i as the latent abundance class for site i , and the probability distribution of N (among sites) as the latent abundance distribution. Note that this definition of N is consistent with the interpretation of the maximum calling index suggested by Knutson et al. (1999) (this is discussed further in *Discussion*). The object of inference is the probability distribution of N given the calling index histories for a sample of sites (i.e., the probabilities $\psi_k = \Pr(N = k)$; $k = 0, 1, 2, 3$, or various functions of these latent class probabilities; see *Inference and prediction*).

The critical statistical consideration in this problem is that the latent abundance class for a site may not be observed due to sampling error. Conditional on the latent abundance class N_i for site i , the observed index y_{it} is multinomial with $N_i + 1$ cells. (recall that $N_i = 0$ is a state). Equivalently, we may say that, given $N_i = k$, y_{it} is a multinomial random variable with four cells and cell probabilities $\pi_{kj} = 0$ if $j > k$. Consequently, we may specify the sequence of multinomial sampling distributions conditional on abundance state, wherein the cell probabilities vary by abundance state according to

	$f(y = 0 N)$	$f(y = 1 N)$	$f(y = 2 N)$	$f(y = 3 N)$
$N = 0$	π_{00}	0	0	0
$N = 1$	π_{10}	π_{11}	0	0
$N = 2$	π_{20}	π_{21}	π_{22}	0
$N = 3$	π_{30}	π_{31}	π_{32}	π_{33}

(1)

where $\pi_{k0} = 1 - \sum_{j=1}^3 \pi_{kj}$. Thus, π_{kk} is the probability the true abundance state is observed and π_{kj} , $j < k$, are

misclassification probabilities. We make the conventional assumption that frogs are not detected where they do not occur so that $\pi_{00} = 1$. In *Alternative parameterization of the conditional sampling distributions*, we present an alternative parameterization of these conditional multinomial sampling distributions that yields a more appealing interpretation of the parameters that is more amenable to parameterizing covariate effects.

The multinomial mixture model is a hierarchical model that yields a joint likelihood for $\{\pi_{jk}\}$ and N_i of the form $[y|N][N]$ where $[y|N]$ is the multinomial sampling distribution conditional on latent abundance, and $[N]$ is the multinomial “prior” distribution of latent abundance. By introducing prior distributions on the remaining parameters, one could jointly estimate the collection of latent abundances, and all model parameters, from the posterior distribution which could be analyzed using Markov chain Monte Carlo (MCMC) methods. However, our main interest is in estimating model parameters (principally the latent abundance probabilities $\{\psi_k\}$), and not the site-specific abundance states. Consequently, we next consider analysis of the marginal likelihood of y , having removed the latent abundance “random effects” by integration.

Focusing on this estimation problem does not preclude estimation of particular values of N (i.e., for a site; we consider frequentist prediction of site-specific latent abundance in *Inference and prediction*). However, for most ecological problems (monitoring programs in particular) that involve sampling large numbers of local populations, interest is primarily focused on structure describing this collection of local populations (i.e., what is usually referred to as the “metapopulation”). In the present case, the metapopulation structure is governed by the ψ_k parameters.

The marginal likelihood

If the set $\{N_i; i = 1, 2, \dots\}$ were known, it would be easy to estimate the detection parameters $\{\pi_{kj}\}$. However, the abundance states are unknown, so the correct site-specific multinomial sampling distribution (which is conditional on N) is not known precisely. Consequently, neither the abundance state probabilities nor the cell probabilities can be estimated simply by tabulating observed index values. Therefore, we focus attention on the integrated likelihood in which each N_i is removed from the conditional multinomial sampling distribution by integration. The integrated likelihood is

$$L(\pi, \psi | y) = \prod_{i=1}^R \left\{ \sum_{k=0}^3 \left[\prod_{j=1}^T f(y_{ij} | N_i = k, \pi_k) \right] \psi_k \right\} \quad (2)$$

where $f(y|N, \pi)$ are specified by Eq. 1, π_k is the k th row of Eq. 1, and $\{\psi_k; k = 0, 1, 2, 3\}$ are the latent abundance class probabilities.

Eq. 2 can be maximized numerically in order to obtain the MLEs of ψ and the detection parameters π . We note that this model may be generalized to enable the

detection parameters, π_{kj} , to depend on time or covariate effects. We consider such generalizations in *Alternative parameterization of the conditional sampling distributions*.

Inference and prediction

We believe that, for most purposes, interest will focus primarily on the latent abundance class distribution $\{\psi_k; k = 0, 1, 2, 3\}$, which is, in effect, a multivariate index to the metapopulation demographic state. That is, the latent abundance class distribution describes the proportion of sites within the sampling frame that are occupied by populations capable of generating index values of 0, 1, 2, or 3. However, several useful (univariate) summaries can be computed from the estimated latent class frequencies. For example, one index that has become very common in many ecological applications is the proportion of sites occupied (“site occupancy”). This is just $1 - \psi_0$. Another possible univariate summary is the average latent abundance class, $\bar{N} = E[N] = \sum_{k=0}^3 k\psi_k$. These are estimated as $1 - \hat{\psi}_0$ and $\sum_{k=0}^3 k\hat{\psi}_k$, respectively.

To obtain variance estimates of these quantities, note that there are $K - 1$ free abundance-class parameters because of the unit sum constraint on ψ_k , and obtain the asymptotic variance–covariance matrix for $\hat{\psi}$ based on the inverse of the Fisher information evaluated at the MLEs, say,

$$\hat{\Sigma} = \widehat{\text{Var}}[(\hat{\psi}_0, \hat{\psi}_1, \hat{\psi}_2)'] = \mathbf{I}^{-1}(\hat{\psi})$$

which is evaluated numerically (e.g., the `nlm` routine in the popular free statistical software R).

For the univariate summaries, note that $\text{Var}(\hat{\psi}_0)$ is the first diagonal element of $\hat{\Sigma}$. To obtain a variance estimate of $\hat{\bar{N}}$ we note that \bar{N} can be expressed as

$$\bar{N} = 3 - 3\psi_0 - 2\psi_1 - \psi_2$$

and hence $\widehat{\text{Var}}(\hat{\bar{N}}) = \mathbf{a}'\hat{\Sigma}\mathbf{a}$ where $\mathbf{a} = (3, 2, 1)$.

In the application to be described in *Application to anuran calling survey data*, the model is parameterized in terms of the multinomial logit of the ψ_k parameters. In this case, application of the delta method is necessary to obtain variance estimates of either the ψ_k parameters or the univariate summaries.

Although the collection of N_i 's has been removed from the likelihood (by integration) to focus estimation on the structural parameters of the prior distribution, it is possible to estimate individual values of N_i if necessary. For this purpose, we adopt the conventional estimated best unbiased predictor (EBUP) which in the present case is

$$\hat{N}_i = E(N_i | \mathbf{y}_i, \hat{\theta}) = \sum_{k=0}^3 k\gamma_i(k)$$

where $\{\gamma_i; k = 0, 1, 2, 3\}$ is the value of $\gamma_i(k) = \Pr(N_i = k | \mathbf{y}_i, \theta)$, evaluated at $\theta = \hat{\theta}$. Here θ indicates the collection of all model parameters $\pi = \{\pi_{kj}\}_{j=k=0}^3$, and

$\psi = (\psi_0, \psi_1, \psi_2, \psi_3)$. Also, $\hat{\theta}$ is an estimate of θ . The probabilities for sample location i , $\{\gamma_i(k)\}_{k=0}^3$, can be computed for any index history given (estimates of) the model parameters, π and ψ . In particular,

$$\Pr(N_i = k | y_i, \theta) \propto \Pr(y_i | N_i, \pi) \Pr(N_i = k).$$

The probabilities $\{\gamma_i(k)\}_{k=0}^3$ constitute an estimate of the posterior predictive distribution of N_i and thus may be used for many standard inference problems regarding N_i , such as point estimation or assessment of uncertainty.

We note that this approach to inference about specific values of the latent abundance class is essentially an empirical Bayes procedure where parameters have been estimated from the marginal likelihood (i.e., having removed the “random effects” by integration) and then used to calculate the conditional posterior of the quantity of interest. We believe that if management interest really were focused on the latent abundance class for a specific site, or complicated functions thereof, one might be better served by adopting a fully Bayesian analysis of the proposed model to more adequately characterize uncertainty due to having estimated the (prior) parameters governing the latent abundance class distribution.

ALTERNATIVE PARAMETERIZATION OF THE CONDITIONAL SAMPLING DISTRIBUTIONS

The fully parameterized model described in *The multinomial mixture model* contains nine parameters (six π 's and three ψ 's). While the parameters of the full model are all identifiable, this model may be an unnecessarily complex description of the detection process. Thus, it may be advantageous to consider more parsimonious models for the conditional multinomial sampling distribution. However, under the parameterization given in *The multinomial mixture model*, it is not at all clear how to define meaningful constraints among the various detection parameters (the π 's of Eq. 1). In addition, there is not a clear and concise parameterization of covariate effects that might influence detection.

To address these issues, we consider the following reparameterization of the multinomial sampling distributions that yields a natural set of reduced submodels and is more amenable to modeling covariate effects:

	$f(y=0 N)$	$f(y=1 N)$	$f(y=2 N)$	$f(y=3 N)$
$N=0$	1	0	0	0
$N=1$	q_1	p_1	0	0
$N=2$	$(1 - \beta_{21})q_2$	$\beta_{21}q_2$	p_2	0
$N=3$	$(1 - \beta_{31})(1 - \beta_{32})q_3$	$\beta_{31}(1 - \beta_{32})q_3$	$\beta_{32}q_3$	p_3

(3)

where $q_k = (1 - p_k)$. The interpretation of these parameters is straightforward. Parameters $\{p_k; k = 1, 2, 3\}$ are “correct classification” probabilities (i.e., the

probability of correctly observing the true abundance class during a sample). Parameters $\{\beta_{kj}\}$ are conditional “misclassification” probabilities: β_{21} is the probability of observing a calling intensity of $y = 1$ given that $N = 2$ and $y \leq 1$, β_{32} is the probability of observing $y = 2$ given that $N = 3$ and $y \leq 2$, and β_{31} is the probability of observing $y = 1$ given $N = 3$ and $y \leq 1$.

This parameterization provides a concise description of the various submodels. For example, one might consider the submodel that assumes $p_1 = p_2 = p_3$, or the various submodels in which only two of the p 's are constrained, e.g., $p_1 = p_2$. Similar constraints can be considered for the β parameters. We do not consider constraints that equate p 's to β 's because these parameters have qualitatively different interpretations. In all, there are 25 possible submodels including the full model containing six parameters.

To keep track of these 25 possible models, it is useful to index them using a parameter index vector that describes the imposed constraints. For example, the model in which $p_1 = p_2$, with p_3 and all β parameters unconstrained has five parameters, and is indicated by the vector (1, 1, 2, 3, 4, 5) with entries indexing constraints among $(p_1, p_2, p_3, \beta_{21}, \beta_{31}, \beta_{32})$. The model with all p 's equal and all β 's equal has two parameters and is denoted by (1, 1, 1, 2, 2, 2). Formally, this vector is just a compact representation of a set of linear constraints among the six detection parameters. Under the full model, denote the (6×1) parameter vector by \mathbf{u} . Denote the $(m \times 1, m < 6)$ reduced parameter vector as \mathbf{v} . Then, the linear constraint can be expressed as $\mathbf{u} = \mathbf{H}\mathbf{v}$ for some $(6 \times m)$ matrix \mathbf{H} with entry (i, j) equal to 1 if the i th element of \mathbf{u} is equal to the j th element of \mathbf{v} and 0 otherwise. The parameter index vector (1, 1, 1, 2, 2, 2) implies that \mathbf{H} is 6×2 with 1 in the first three rows of column 1 and in the last three rows of column 2.

Modeling covariate effects on the detection process

In many applications, information on one or more covariates that affect the detection of frogs will be collected. For example, ambient temperature is collected during many anuran surveys (including NAAMP). Because calling activity of frogs is affected by temperature (Johnson and Batie 2001), it is important to accommodate this information in the detection probability parameters. Also, variation in detection probability over time may occur as a result of environmental conditions that are not quantified. Because there may be as many as six detection probability parameters, there are many ways of parameterizing covariate effects. It would be possible to parameterize the covariate effects on all of the parameters (the p 's and β 's), but this model becomes unwieldy.

Alternatively, a sensible and concise approach is to model covariates as having additive effects on the logits of p_1 , p_2 , and p_3 but not on the β parameters. Thus, it is assumed that the effect of covariates on π_{kj} (Eq.

TABLE 1. Summary of models fit to green frog data.

Model index vector						np	AIC	$\hat{\psi}_0$	$\hat{\psi}_1$	$\hat{\psi}_2$	$\hat{\psi}_3$	\hat{b}_1	\hat{b}_3
p_1	p_2	p_3	β_{21}	β_{31}	β_{32}								
1	2	2	3	3	3	8	605.69	0.47	0.33	0.15	0.04	−3.48	2.59
1	2	2	3	3	4	9	606.06	0.47	0.33	0.15	0.05	−3.73	2.62
1	2	2	3	4	4	9	607.07	0.47	0.33	0.15	0.05	−3.85	2.63
1	2	1	3	3	3	8	607.10	0.45	0.36	0.16	0.03	−3.54	2.42
1	1	1	2	2	2	7	607.19	0.42	0.42	0.13	0.04	−3.29	2.15
1	1	1	2	2	3	6	712.04	0.12	0.19	0.55	0.14		

Notes: Rows 1–5 are the top five models (by Akaike's Information Criteria [AIC]) containing time effects (parameters b_1 and b_3) in detection probability. Row 6 is the best model without time effects. An empty cell indicates that the model does not contain that parameter; np = number of model parameters.

1) is solely through the p 's. So, for example, any effect on p_3 has the opposite effect on the marginal detection probability $\pi_{32} = \Pr(y = 2 | N = 3)$. That is, since $\Pr(y = 2 | N = 3) = \pi_{32}(1 - p_3)$, a factor that has a positive influence on p_3 must have a negative influence on the marginal probability of observing $y = 2$ given that $N = 3$.

Let $\{p_k^i; k = 1, 2, 3\}$ be the correct classification probability for site i during sample t given that $N_i = k$, $p_k^i = \Pr(y_{it} = k | N_i = k)$. We consider additive models of the form

$$\text{logit}(p_k^i) = \alpha_k + \sum_{l=1}^L b_l x_{l,it} \quad (4)$$

where $\{x_{l,it}; l = 1, 2, \dots, L\}$ is the value of the l th covariate for observation it . The indices of the parameter index vector described in *Alternative parameterization of the conditional sampling distributions* now index the α_k parameters instead of p_k . For each of those 25 models, the l th covariate may be included in the model, or not, so that there are 25×2^L possible models with L covariates.

APPLICATION TO ANURAN CALLING SURVEY DATA

Here, we consider data collected in Maryland during 2001 by the North American Amphibian Monitoring Program (NAAMP). The NAAMP is a roadside survey modeled after the North American Breeding Bird Survey (BBS; Robbins et al. 1986). Sample sites consist of roadside “stops” near wetlands where observers record the calling index of every species detected. A detailed description of the protocol is given by Weir and Mossman (2005). Some relevant discussion of protocol and implementation for specific programs can be found in Shirose et al. (1997), Crouch and Paton (2002), and Genet and Sargent (2003).

Data analyzed here are index values of green frog (*Rana clamitans*) calling activity recorded at 220 NAAMP sample locations. These data and a computer program to fit the multinomial mixture models have been provided in the Supplement. The three NAAMP samples were collected during well-defined NAAMP sampling “windows” designed to detect all species (Weir and Mossman 2005). Green frogs breed later in

the season than many species; in Maryland, their peak breeding activity (and hence, greatest vocalization) occurs during the third sampling window (1–30 June). Because of this, we might expect temporal variation in detection probability to be an important source of variation in the observed index data. We therefore considered a model in which detection probability varied among the three sample periods by parameterizing temporal variation as an additive effect on the logit scale. That is,

$$\begin{aligned} \text{logit}(p_k^{i1}) &= \alpha_k + b_1 \\ \text{logit}(p_k^{i2}) &= \alpha_k \\ \text{logit}(p_k^{i3}) &= \alpha_k + b_3 \end{aligned} \quad (5)$$

where $\{p_{k,it}; k = 1, 2, 3\}$ are described in *Modeling covariate effects on the detection process*.

We fit the suite of 25 detection probability models representing constraints among the various detection parameters (described in *Alternative parameterization of the conditional sampling distributions*) to the green frog data twice: once with and once without time effects. The five models having the best (i.e., smallest) Akaike Information Criterion score (AIC; Burnham and Anderson 1998) are given in Table 1 (first five rows). Note that all of these top models contain temporal variation in detection probability. For comparison, the best model without temporal variation in detection probability is also given (row 6). The AIC of this model was higher than all 25 of those models that allowed for temporal variation.

Among those models with time effects, there is little difference in the estimated latent abundance distribution. From these results, it is estimated that approximately 47% of sites are “unoccupied” by green frogs, 4% have abundance levels capable of generating an index value of $y = 3$, and the remaining 48% are of intermediate levels of abundance (33% have $N = 1$ and 15% have $N = 2$). To give some context to these estimates, we note that the observed frequency (proportion) of maximum index values at the 220 sites were 134 (0.609), 62 (0.282), 19 (0.086), and 5 (0.023) for $\max_i(y_{it}) = 0, 1, 2, 3$, respectively. Thus, the effect of accounting for detectability adjusts the abundance dis-

TABLE 2. Summary of models considering temperature effects fit to green frog data.

Model index vector						np	AIC	$\hat{\psi}_0$	$\hat{\psi}_1$	$\hat{\psi}_2$	$\hat{\psi}_3$	\hat{b}_1	\hat{b}_3	\hat{c}_1	\hat{c}_2
p_1	p_2	p_3	β_{21}	β_{31}	β_{32}										
1	2	2	3	3	3	9	591.95	0.47	0.33	0.15	0.04	-2.42	2.04	0.13	
1	2	2	3	3	4	10	592.45	0.47	0.33	0.15	0.05	-2.46	2.08	0.12	
1	2	2	3	3	3	10	592.56	0.47	0.33	0.16	0.04	-2.20	2.13	0.20	-0.01
1	2	2	3	3	4	11	592.80	0.48	0.32	0.16	0.05	-3.74	2.17	0.24	-0.02
1	2	2	3	3	3	8	595.27	0.46	0.33	0.16	0.05	-3.47	2.58		
1	2	2	3	3	4	9	595.56	0.46	0.33	0.16	0.05	-3.74	2.61		
1	2	2	3	3	3	8	617.96	0.38	0.40	0.16	0.05			0.34	-0.01
1	1	1	2	2	2	7	617.97	0.34	0.47	0.14	0.04			0.32	-0.01
1	2	2	3	3	3	7	618.46	0.40	0.39	0.16	0.05			0.28	
1	1	1	2	2	2	6	618.69	0.37	0.45	0.13	0.04			0.26	

Notes: The best two models (by AIC) for each possible set of effects are given; model sets are separated by blank rows. An empty cell indicates that the model does not contain that parameter; np = number of model parameters.

tribution upward a moderate amount, as expected, to reflect that green frogs likely occur at some sites where non-detection occurred, and occur at higher levels of abundance than observed at other sites.

Note that the estimated time effects (\hat{b}_1 and \hat{b}_3 in Table 1) (parameterized as differences from period 2 detection) indicate that detection probability increases from period 1 to period 3, consistent with the notion that green frog breeding activity occurs later in the season. There is a large difference in estimated latent abundance distribution when the time effects are omitted from the model (i.e., compare row 6 to rows 1–5 of Table 1), suggesting the importance of properly modeling the detection process. The model without time effects understates detection probabilities, and, hence, overstates abundance levels.

Fitted values of the multinomial cell probabilities (Eq. 1), conditional on N , were calculated under the best model (row 1 in Table 1) where $p_2 = p_3$ and $\beta_{21} = \beta_{31} = \beta_{32}$. Estimated cell probabilities for period 2 (the baseline period, containing a zero time effect) are

	$f(y=0 N)$	$f(y=1 N)$	$f(y=2 N)$	$f(y=3 N)$
$N=0$	1.000	0	0	0
$N=1$	0.758	0.242	0	0
$N=2$	0.772	0.149	0.078	0
$N=3$	0.647	0.125	0.149	0.078

We note a high probability of observing zero under any abundance state and considerably higher probability of correctly detecting $N = 1$ than either of $N = 2$ or $N = 3$.

Detection response to temperature

Finally, we considered the addition of temperature as a covariate effect on detection probability. We considered linear and quadratic temperature effects as additive effects in Eq. 5, yielding models of the form

$$\text{logit}(p_k^{i1}) = \alpha_k + b_1 + c_1 \text{temp}_{i1} + c_2 \text{temp}_{i1}^2$$

$$\text{logit}(p_k^{i2}) = \alpha_k + 0 + c_1 \text{temp}_{i2} + c_2 \text{temp}_{i2}^2$$

$$\text{logit}(p_k^{i3}) = \alpha_k + b_3 + c_1 \text{temp}_{i3} + c_2 \text{temp}_{i3}^2 \quad (6)$$

where “temp_{*it*}” is the observed temperature at site *i* during sample *t*. Note that we have deviated somewhat from the generic representation for the model presented in Eq. 4 so as to distinguish time effects (parameters b_1 and b_3) from temperature effects (c_1 and c_2). For these analyses, we omitted eight observations with missing temperature data; for the purposes of comparison, we refit the best models of Table 1 using the modified data set (using the reduced data set, the AIC scores were 595.27 and 595.56 for the best two models summarized in Table 1). Because temporal variation may be an artifact of seasonal variation in temperature, we considered models with temperature, but without time effects. This yields the following five model sets (containing 25 models each) varying by covariate effects included in the model:

- 1) time effects only (i.e., Table 1);
- 2) temperature (linear), no time effects;
- 3) temperature (linear, quadratic), no time effects;
- 4) time + temperature (linear);
- 5) time + temperature (linear, quadratic).

The results for the top two models for each set (based on AIC) are summarized in Table 2. Each pair of rows (separated by a horizontal line) pertains to a particular model set that can be deduced by the presence or absence of a particular parameter estimate. Row pairs are ordered by AIC of the best model within the class. Thus, we see that the model containing time effects and a linear temperature effect (rows 1 and 2 in Table 2) is favored over the other models, but just slightly over the model with time and a quadratic temperature effect (rows 3 and 4 in Table 2). Note (column \hat{c}_1 in

Table 2) that the positive (linear) temperature effect indicates increasing detection probability with temperature and the negative quadratic indicates that detection probability is concave, obtaining a maximum detection probability at 17.3 (°C) for the best model in that set. Note also that rows 5 and 6 in Table 2 are the best two models in Table 1, with slight differences reflecting that eight observations without temperature data have been omitted.

Inclusion of a temperature effect in the model does moderate the time effects to some extent (reducing their magnitude), suggesting that some of the temporal variation in detectability is due to temperature. Finally, the estimated abundance-class frequencies differ very little among these models, even for models that contain a temperature effect, but no time effect. Under the best model in Table 2, we computed the two univariate summaries of the latent class distribution described in *Inference and prediction*. The estimated probability of occurrence (i.e., $\Pr[N > 0]$) under this model is $1 - \hat{\psi}_0 = 0.53$ (SE = 0.0433). The estimated mean latent class is $\hat{N} = 0.77$ (SE = 0.074).

DISCUSSION

Despite the prevalence of monitoring programs for amphibians, and anurans in particular, there has been little development of statistical models that account for imperfect detection in calling surveys. Models for estimating “site occupancy” (probability of occurrence) based on observed presence/absence (e.g., MacKenzie et al. 2002) enable formal modeling of detection probability. However, that approach enables only a crude characterization of abundance state by classifying sites as being either occupied or unoccupied. Changes in anuran abundance can go undetected when abundance index data are reduced to simple detection/nondetection; a two-state population description (occupied or unoccupied) is not sensitive to underlying population shifts that may not affect site occupancy but that may be important for conservation or management objectives.

Here, we have developed a general model where the sampling distribution of the observed index data for a site is defined conditional on the “latent abundance class” for that site. The model allows this latent abundance class to be unobserved due to sampling error. Importantly, the model enables a multinomial summary of the observed index data to be obtained (from which various univariate summaries may be obtained, including probability of occurrence). The latent abundance class is interpreted according to the maximum index value that the population at a given site could generate. For example, a site has an abundance class of $N = 3$ if it contains a population capable of yielding an index value of $y = 3$. The proposed framework enables the construction of very general models of the detection process and yields estimates of the underlying latent abundance class distribution that are adjusted for sam-

pling conditions that can vary spatially, temporally, and as a function of many controllable and uncontrollable factors.

Although our definition of abundance class is not precisely linked to population size, it is easily interpretable in terms of the calling index and in the context of the sampling process that generated the data. Our definition of abundance class also is consistent with contemporary notions of characterizing metapopulation structure (e.g., “site occupancy”). And, the interpretation of the latent abundance class is consistent with the interpretation of the maximum calling index (at a site) forwarded by Knutson et al. (1999). They provide an analysis of similar data but reduced the observed index data to the maximum observed index value at that site arguing that: “The maximum value represents the highest population level an individual survey location could produce for a given species—the wetland at its best.” (Knutson et al. 1999:1439) However, they do not account for the possibility that this maximum attainable index value may not have been observed during sampling.

As with any animal monitoring program, there is some debate over the best way in which to collect abundance information, and there is debate about the utility of information provided by anuran surveys based on calling indices. However, sampling based on calling surveys is the only practical method for monitoring anurans over large scales. We note that we are not advocates of calling indices to the exclusion of other types of data, and our model does not impose a precise link between observed calling indices and actual abundance. While several studies have asserted a linkage between calling indices and abundance (Mossman et al. 1998, Knutson et al. 1999, Stevens et al. 2002), to our knowledge, the only published study that evaluates this linkage directly is Shirose et al. (1997). They noted a significant linear relationship between number of individuals perceived to be calling and the number of individuals captured using intensive survey techniques for several species. Although they attempted to count the number of unique vocalizing individuals, the categorical index collected by most anuran monitoring programs represents a coarse summary of such counts. The relationship between calling intensity and abundance is a topic of some importance that will undoubtedly be the focus of future research.

We applied the proposed model to green frog index data collected in 2001 on NAAMP routes in Maryland. We propose the estimated latent abundance distribution for monitoring and assessment purposes because it yields a summary metric that is adjusted for variation in detection probability induced by variation in factors which influence the detectability of anurans (e.g., temperature). The estimated abundance distribution was shifted upwards from the observed calling index frequencies to reflect the indicated structure in detection probability. Since factors that influence detectability

will vary spatially, temporally, and in response to environmental conditions, it is necessary to account for detectability in order to properly partition spatial and temporal variation in the observed data into variation due to ecological processes of interest (i.e., abundance) and variation due to sampling processes.

ACKNOWLEDGMENTS

We thank an anonymous referee and Scott Harrison for many helpful comments on the manuscript. We also thank Linda Weir for providing the data used in our analyses.

LITERATURE CITED

- Alford, R. A., and S. J. Richards. 1999. Global amphibian declines: a problem in applied ecology. *Annual Review of Ecology and Systematics* **30**:133–165.
- Bishop, C. A., K. E. Petit, M. J. Gartshore, and D. A. MacLeod. 1997. Extensive monitoring of anuran populations using call counts and road transects in Ontario (1992 to 1993). Pages 149–160 in D. M. Green, editor. *Amphibians in decline: Canadian studies of a global problem*. Society for the Study of Amphibians and Reptiles, St. Louis, Missouri, USA.
- Bridges, A. S., and M. E. Dorcas. 2000. Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. *Copeia* **2000**:587–592.
- Burnham, K. P., and D. R. Anderson. 1998. *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New York, New York, USA.
- Carroll, R. J., and F. Lombard. 1985. A note on N estimators for the binomial distribution. *Journal of the American Statistical Association* **80**:423–426.
- Crouch, W. B., III, and P. W. Paton. 2002. Assessing the use of call surveys to monitor breeding anurans in Rhode Island. *Journal of Herpetology* **36**:185–192.
- Genet, K. S., and L. G. Sargent. 2003. Evaluation of methods and data quality from a volunteer based amphibian call survey. *Wildlife Society Bulletin* **31**:703–714.
- Hall, R. J., and C. A. Langtimm. 2001. The U.S. National Amphibian Research and Monitoring Initiative and the role of protected areas. *George Wright Forum* **18**:14–25.
- Houlahan, J. E., C. S. Findlay, B. R. Schmidt, A. H. Meyer, and S. L. Kuzmin. 2000. Quantitative evidence for global amphibian population declines. *Nature* **404**:752–755.
- Johnson, D. H., and R. D. Batie. 2001. Surveys of calling amphibians in North Dakota. *Prairie Naturalist* **33**:227–247.
- Knutson, M. G., J. R. Sauer, D. A. Olsen, M. J. Mossman, L. M. Hemesath, and M. J. Lannoo. 1999. Effects of landscape composition and wetland fragmentation on frog and toad abundance and species richness in Iowa and Wisconsin, U.S.A. *Conservation Biology* **13**:1437–1446.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**:2248–2255.
- Mossman, M. J., L. M. Hartman, R. Hay, J. R. Sauer, and B. J. Dhuey. 1998. Monitoring long term trends in Wisconsin frog and toad populations. Pages 169–198 in M. J. Lannoo, editor. *Status and conservation of midwestern Amphibians*. University of Iowa Press, Iowa City, Iowa, USA.
- Oseen, K. L., and R. J. Wassersug. 2002. Environmental factors influencing calling in sympatric anurans. *Oecologia* **133**:616–625.
- Reaser, J. K. 2000. *Amphibian declines: an issue overview*. Federal Taskforce on Amphibian Declines and Deformities, Washington, D.C., USA.
- Robbins, C. S., D. A. Bystrak, and P. H. Geissler. 1986. *The Breeding Bird Survey: its first fifteen years, 1965–1979*. Resource Publication 157. U.S. Department of Interior, Fish and Wildlife Service, Washington, D.C., USA.
- Royle, J. A. 2004a. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**:108–115.
- Royle, J. A. 2004b. Modeling abundance index data from anuran surveys. *Conservation Biology* **18**:1378–1385.
- Shirose, L. J., C. A. Bishop, D. M. Green, C. MacDonald, R. J. Brooks, and N. J. Helferty. 1997. Validation tests of an amphibian call count survey technique in Ontario, Canada. *Herpetologica* **53**:312–320.
- Stevens, C. E., A. W. Diamond, and T. S. Gabor. 2002. Anuran call surveys on small wetlands in Prince Edward Island, Canada restored by dredging of sediments. *Wetlands* **22**:90–99.
- Weir, L. A., and M. J. Mossman. 2005. North American Amphibian Monitoring Program. Pages 307–313 in M. J. Lannoo, editor. *Declining amphibians: A United States response to the global phenomenon*. University of California Press, Berkeley, California, USA.

SUPPLEMENT

Data and computer programs for fitting multinomial mixture models used in the main article are available in ESA's Electronic Data Archive: *Ecological Archives* E086-132-S1.